三维片上网络正四面体裂变拓扑结构研究 *

郑亚振，张大坤

(天津工业大学 计算机科学与软件学院，天津 300387)

摘　要：旨在研究新型三维片上网络正四面体裂变拓扑结构，给出了该拓扑结构的生成过程；对该拓扑结构进行了编码设计和路由设计。通过对 gpNoCsim 片上网络仿真器进行三维扩展，对正四面体裂变拓扑结构进行性能仿真实验。仿真结果表明，在均匀负载模式下，正四面体裂变拓扑结构的平均延时和平均跳数均低于 Mesh 结构，当注入率为 0.02 时，平均延时比 Mesh 结构低 16.8%、平均跳数比 Mesh 结构少 5.5%；在局部负载模式下，当注入率大于 0.008 时，正四面体裂变拓扑结构的平均延时和平均跳数与 Mesh 结构相比，均有明显改善；当注入率为 0.014 时，平均延时比 Mesh 结构降低 18.7%、平均跳数比 Mesh 结构减少 9.6%。说明正四面体裂变拓扑结构可用于三维片上网络拓扑结构设计。

关键词：三维片上网络；正四面体裂变；拓扑结构；平均延时；平均跳数

中图分类号：TP39　　doi: 10.3969/j.issn.1001-3695.2017.08.0888

Research on tetrahedron fission topology in three dimensional network-on-chip

Zheng Yazhen, Zhang Dakun

(*School of Computer Science & Software Engineering*, *Tianjin Polytechnic University*, *Tianjin* 300387)

Abstract: This paper aims to research the tetrahedron fission topology that is a new type of three dimensional Network-on-Chip topology, gives generation process of topology and the design of encoding and routing. Through a three dimensional extension of the gpNoCsim which is a simulator on chip network, the simulation experiment of the tetrahedron fission topology is conducted. The simulation results show, tetrahedron fission topology can achieve lower average latency and fewer average hops than Mesh under uniform traffic pattern, When the injection rate is 0.02, comparing with Mesh, the average latency degrades by 16.8 % and the average hops decrease 5.5% ; Under localized traffic pattern, the average delay of tetrahedral fission topology and the average hops are significantly improved when the injection rate is greater than 0.008, Comparing with Mesh , tetrahedron fission leads to 18.7% decrease of the average latency and 9.6% decrease of the average hops when the injection rate is 0.014. It is shown that the tetrahedron fission topology can be used in the design of three dimensional Network-on-Chip topology.

Key Words: three-dimensional network on chip; tetrahedron fission; topology; average latency; average hops

0　引言

随着大规模集成电路技术的发展，片上系统(system-on-chip,SoC)和二维片上网络(two dimensional network-on-chip,2D NoC)相继产生。2D NoC 规模逐渐增大，使 2D NoC 在面积、功耗、布局布线以及封装密度等方面都达到了瓶颈[1]，因而产生了三维片上网络(three dimensional net-work-on-chip,3D NoC)，并已成为研究热点[2~5]。在 3D NoC 的研究中，拓扑结构是关键问题之一[6]；3D NoC 拓扑结构包括规则拓扑和非规则拓扑两大类，在经典规则拓扑结构基础上进行性能分析的研究较多，如 3D Mesh [7]结构和 3D Torus [8]结构等，但关于新型非规则 3D NoC 拓扑结构的研究较少。与规则拓扑结构相比，非规则拓扑结构可以面向不同的领域，根据特定的应用需求而定制、具有很好的应用前景，这方面的研究已取得了一些进展。如芬兰图尔库大学 Yin 等人提出的蜂窝状 3D NoC 拓扑结构[9]；湖南大学贺旭等人提出的三维超立方体片上网络拓扑结构[10]；西安电子科技大学刘有耀等人对 NoC 拓扑结构与通信方法进行了研究，对典型的 3D NoC 架构进行了总结，并提出了 3 种 3D NoC 的架构，包括提出了三维超立方体双环架构、三维 Torus 连接的 Petersen 图架构、三维长方形扭花环网格架构[11]。本文旨在研究新型非规则 3D NoC 拓扑结构，提出了正四面体裂变拓扑结构（选题源于国家自然科学基金课题：61272006）。该拓扑结构具有节点度低(恒为 3)、对称性好、路由简单等优点，并使网络扩展从“加法”转为“裂变”，可以用于个性化 3D NoC 拓扑结构设计，研究具有探索性和创新性。

基金项目：国家自然科学基金面上项目（61272006）

作者简介：郑亚振（1989-），男，河南周口人，硕士研究生，主要研究方向为三维片上网络及拓扑结构（zhangdakun2013@163.com）；张大坤（1960-），女，教授，博导，博士，主要研究方向为三维片上网络、大数据可视分析、虚拟现实技术等．

1　正四面体裂变拓扑结构及路由设计

1.1　正四面体裂变拓扑结构生成

正四面体是柏拉图立体之一[12]，正四面体裂变拓扑结构是由正四面体"裂变"而来，带标注的正四面体模型如图 1 所示。正四面体节点一级裂变拓扑过程为：从图 1 任选一个节点 No_0，从 No_0 出发沿与之相连的 3 条棱移动一段距离后生成 3 个新节点 No_{00}、No_{01}、No_{02}，然后依次相连这 3 个节点（把互连的 3 个节点称为裂变簇）并删除 No_0，这样就完成了 No_0 的裂变，No_0 裂变过程如图 2 所示。图中圆圈表示网络节点。一般的网络节点中包含一个路由器和若干个 IP 核，IP 核可以是处理器、存储器或具有其他功能的器件，IP 核通过本地端口和路由器相连。

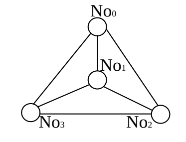


图 1　正四面体标注

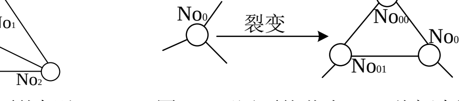
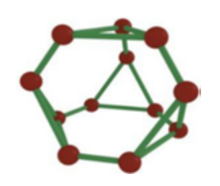


图 2　正四面体节点 No_0 裂变过程

按照上述裂变过程依次完成图 1 中各个节点的裂变和互连，就得到了正四面体裂变拓扑结构。在该结构中有 12 个网络节点，每个节点的度都是 3，有 4 个裂变簇，用 A、B、C 和 D 来表示。正四面体裂变拓扑结构球棍模型如图 3 所示。


图 3　正四面体裂变拓扑结构球棍模型

1.2　正四面体裂变拓扑结构节点编码与路由

1）节点编码

合理的节点编码方案能够简化路由协议的复杂度、提高网络性能、降低延时。正四面体裂变拓扑结构是由正四面体裂变而来，为了方便寻址需要给裂变簇 A、B、C 和 D 进行编码，该结构中共有 12 个网络节点需要 4 bit 来编码，各个网络节点的编码是由裂变簇通过向左移动两位加上节点在裂变簇内的编号而生成，具体编码如表 1 所示。

表 1　正四面体裂变结构各节点及裂变簇编码

正四面体节点	裂变簇	编码
No_0 节点	No_0 节点经过裂变后变成裂变簇 A，裂变簇中有 No_{00}、No_{01}、No_{02} 三个节点	裂变簇 A 编码为 00，其内部节点编码为 No_{00}：0000，No_{01}：0001，No_{02}：0010
No_1 节点	No_1 节点经过裂变后变成裂变簇 B，裂变簇中有 No_{10}、No_{11}、No_{12} 三个节点	裂变簇 B 编码为 01，其内部节点编码为 No_{10}：0100，No_{11}：0101，No_{12}：0110
No_2 节点	No_2 节点经过裂变后变成裂变簇 C，裂变簇中有 No_{20}、No_{21}、No_{22} 三个节点	裂变簇 C 编码为 10，其内部节点编码为 No_{20}：1000，No_{21}：1001，No_{21}：1010
No_3 节点	No_3 节点经过裂变后变成裂变簇 D，裂变簇中有 No_{30}、No_{31}、No_{32} 三个节点	裂变簇 D 编码为 11，其内部节点编码为 No_{30}：1100，No_{31}：1101，No_{32}：1110

2）路由器结构设计

在正四面体裂变拓扑结构中，每个网络节点中的器有 3 个与相邻网络节点连接的端口和若干个本地端口，本设计中每个节点搭载 4 个 IP 核，故需要 4 个本地端口，本地端口的一端与路由器相连，另一端与本地 IP 核相连。每个端口包括输入通道和输出通道，每个通道可以从相连接的路由器或者 IP 核接收数据包或者发送数据包。其中每个物理通道又被划分为多个虚通道，一般被划分为 2 个、4 个或 8 个，本设计选取 4 个，以便有效地避免死锁。路由器中包括路由逻辑、交换开关和虚通道仲裁单元等，通过路由计算把网络节点接收的数据包发送到正确的输出通道上，物理通道被各个虚通道采用轮转法使用。

3）路由算法设计

正四面体裂变拓扑结构中有 12 个网络节点、4 个裂变簇。表 1 为正四面体裂变拓扑结构的编码具体情况。由于本结构具有分级的特征，故采用分级路由的思想，其中 *dest* 代表目的节点，*curr* 代表当前节点，*a* 值判断数据包是否到达 *dest* 所在的裂变簇，*result* 值判断数据包是否到达 *dest* 节点。具体的路由过程如下：

*a)*数据包中包含有目的节点地址 *dest*，当正四面体裂变拓扑结构节点 *curr* 收到一个数据包时，计算 $result= curr \oplus dest$，其中 ⊕ 为异或运算，转向步骤 *b)*；

*b)*判断 *result* 值，若 *result* 值不为 0，则转向步骤 *Step3*，否则转向步骤 *f)*；

c)计算 currA=curr/4，destA= dest/4，currB= curr mod 4，destB= dest mod 4，a= currA ⊕ destA，并判断 a 的值，若 a 值不为 0，则转向步骤 d)，否则转向步骤 e)；

*d)*判断 *currA*、*destA*、*destB* 的值将数据包发送到相邻的裂变簇或者发送到距目的裂变簇最近的节点并将其节点更新为当前节点 *curr*，转向步骤 *a)*；

e）目的节点和当前节点在同个裂变簇，然后判断 *currB* 和 *destB* 的值，将数据包发送到相应的节点并将其节点更新为当前节点 *curr*，转向步骤 *a)*；

f）数据包已经到达目的节点，则将数据包发送到对应的 IP 核。

2　正四面体裂变拓扑结构性能仿真

2.1　仿真实验设计

1）仿真平台与实验方案

本实验采用美国切斯特大学 Hossain 等人研发的 gpNoCsim[13]片上网络仿真器，对该仿真器进行了三维扩展，实现了正四面体裂变拓扑结构的编码。将正四面体裂变拓扑与 2D Mesh 结构在同等网络规模下进行仿真实验，比较两种结构的吞吐量、平均延时和平均跳数等性能指标，以验证正四面体裂变拓扑结构的优越性。正四面体裂变拓扑结构共有 12 个网络节点，故 2D Mesh 网络规模大小需要设置为 4×3，共 12 个网络节点。每个网络节点中的路由器通过本地端口与 4 个 IP 核相连，故两种结构搭载的 IP 核数都是 48 个。实验中，2D Mesh 采用 XY 路由算法，正四面体裂变拓扑结构变采用本文 1.2 中所设计的路由算法。实验采用局部负载模式和均匀负载模式进行仿真。均匀负载模式下，网络中的流量分布均匀，每个节点收到数据包的概率相等；局部负载模式下，70%的流量限制在簇内 4 个 IP 核中，30%的流量随机地发送到簇外的其他节点。

2）参数配置

对 gpNoCsim 仿真器所有需要的参数进行了设置，其中主要的网络参数配置如下：每个物理通道中虚通道数设置为 4，每个虚通道的大小设置为存储 4 个 flit，消息长度设置为 200 字节，微片长度设置为 150 位， 仿真实验每次运行 20000 个时钟周期，其中预热期设置为前 10%的时钟。为了使仿真结果更为精确，仿真器在相同的参数配置下重复运行 20 次，将算术平均值作为仿真结果输出值。

2.2　吞吐量对比分析

1）均匀负载模式吞吐量对比分析

通过仿真实验在均匀负载模式下，得到了正四面体裂变拓扑结构和 2D Mesh 结构网络吞吐量的实验数据。分析实验室数据可知，当注入率小于 0.016 时，此时网络未发生拥塞，两者在相同的注入率下，网络吞吐量基本一致；随着数据包注入率的继续增加，网络逐渐变得繁忙，正四面体裂变拓扑结构将达到饱和状态。对于正四面体裂变拓扑结构，当注入率到达 0.018 时，网络达到饱和状态，饱和时的吞吐量为 0.17；对于 2D Mesh 结构，当正四面体裂变拓扑结构网络达到饱和时，2D Mesh 结构网络的吞吐量为 0.18；正四面体裂变拓扑结构饱和时的吞吐量比 2D Mesh 结构低 5.6%。两种拓扑结构吞吐量与注入率的关系曲线如图 4 所示。

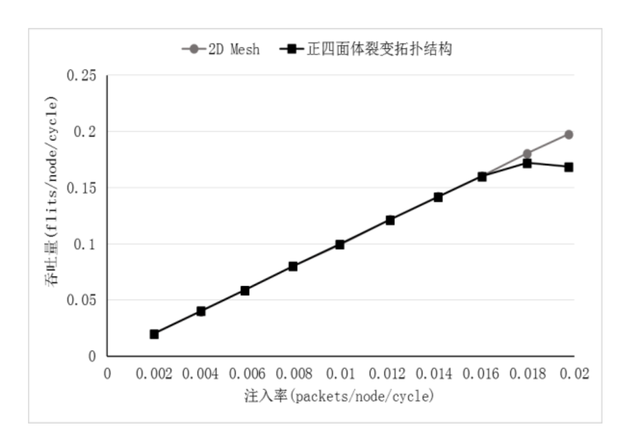


图 4　均匀负载模式下两种拓扑结构吞吐量的对比

2）局部负载模式吞吐量对比分析

在局部负载模式下，对正四面体裂变拓扑结构和 2D Mesh 结构吞吐量实验数据进行分析，当注入率小于 0.004 时，两种拓扑结构的网络吞吐量基本一致；随着数据包注入率的增加，网络流量也随之增大，两种拓扑结构相继达到饱和状态。对于正四面体裂变拓扑结构，当注入率增加到 0.008 时，其吞吐量为 0.39，网络达到饱和状态；对于 2D Mesh 结构，当注入率增加到 0.008 时，其吞吐量为 0.41，网络达到饱和状态；在饱和状态下，正四面体裂变拓扑结构的吞吐量比 2D Mesh 结构低 4.9%；但当网络达到饱和后随着数据包注入率的增加，与正四面体裂变拓扑结构相比，2D Mesh 结构的吞吐量明显下降，在注入率为 0.01 时，正四面体裂变拓扑结构的吞吐量比 2D Mesh 结构高 6.1%；在注入率为 0.012 时，正四面体裂变拓扑结构的吞吐量比 2D Mesh 结构高 16.7%。两种拓扑结构在局部负载模式下吞吐量与注入率的关系曲线如图 5 所示。

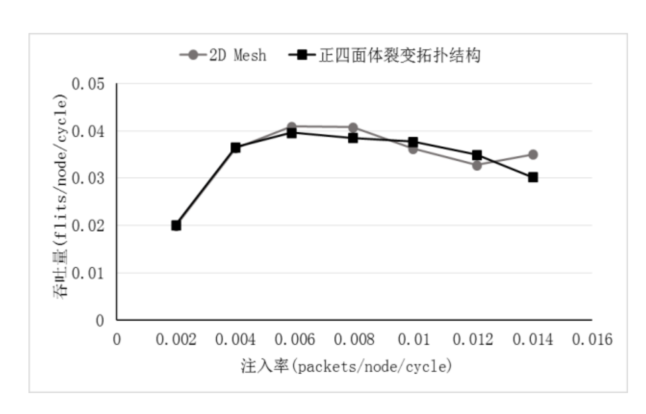


图 5　局部负载模式下两种拓扑结构吞吐量的对比

2.3　平均延时对比分析

1）均匀负载模式平均延时对比分析

实验室数据表明在均匀负载模式、不同注入率下，正四面体裂变拓扑结构的平均延时均优于比 2D Mesh 结构，但随着数据包注入率的增加，两者的平均延时也随之增加。在注入率为 0.016 时，正四面体裂变拓扑结构的平均延时比 2D Mesh 结构低 7.2%；当注入率为 0.018 时，正四面体裂变拓扑结构的平均

延时比 2D Mesh 结构低 11.0%；当注入率为 0.02 时，正四面体裂变拓扑结构的平均延时比 2D Mesh 结构低 16.8%。两种拓扑结构在均匀负载模式下平均延时与注入率的关系曲线如图 6 所示。

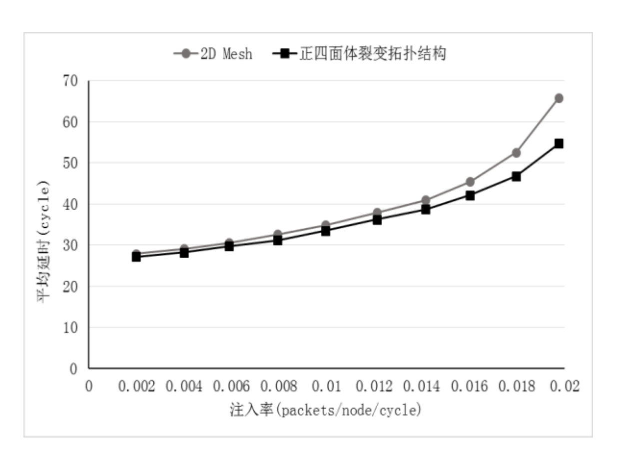


图 6　均匀负载模式下两种拓扑结构平均延时的对比

2）局部负载模式平均延时对比分析

在局部负载模式下、注入率较小时，正四面体裂变拓扑结构和 2D Mesh 结构的平均延时均较低；随着数据包注入率的逐渐增加，正四面体裂变拓扑结构和 2D Mesh 结构的平均延时也随之增加；当注入率小于 0.008 时，正四面体裂变拓扑结构的平均延时增长幅度略大于 2D Mesh 结构；当注入率大于 0.008 时，与正四面体裂变拓扑结构相比 2D Mesh 结构的平均延时增长幅度较大，这是由于 2D Mesh 结构最外侧节点的节点度较小，当注入率增加时，这些节点更容易发生阻塞，所以处理数据包的等待时间明显增加；当注入率较高时，与均匀负载模式相比，两种拓扑结构在局部负载模式下的平均延时更大，这是由于网络阻塞导致增加的时间要远大于由于平均跳数缩短而减少的时间；注入率为 0.01 时，正四面体裂变拓扑结构的平均延时比 2D Mesh 结构低 19.1%；注入率为 0.014 时，正四面体裂变拓扑结构的平均延时比 2D Mesh 结构低 18.7%。两种拓扑结构在局部负载模式下平均延时与注入率的关系曲线如图 7 所示。

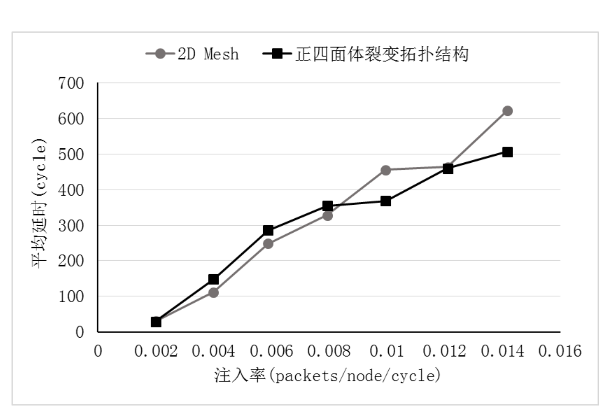


图 7　局部负载模式下两种拓扑结构平均延时的对比

2.4　平均跳数对比分析

1）均匀负载模式下平均跳数对比分析

在均匀负载模式、不同的注入率下，2D Mesh 结构和正四面体裂变拓扑结构收到数据包的总数、平均跳数的实验数据如表 2 所示，由实验数据可知，该负载模式下正四面体裂变拓扑结构的平均跳数比 2D Mesh 结构大约减少 5.5%。

表 2　均匀负载模式下两种拓扑结构平均跳数的对比

注入率	NoC 结构			
	2D mesh		正四面体裂变结构	
	收包总数	平均跳数	收包总数	平均跳数
0.002	1915.30	4.1807	1909.00	3.9651
0.004	3835.20	4.1889	3840.70	3.9547
0.006	5626.80	4.1815	5639.90	3.9567
0.008	7665.80	4.1836	7686.85	3.9607
0.010	9537.85	4.1847	9569.50	3.9573
0.012	11668.60	4.1795	11637.35	3.9579
0.014	13626.65	4.1856	13600.05	3.9582
0.016	15378.55	4.1858	15347.10	3.9568
0.018	17329.90	4.1844	16515.30	3.9575

2）局部负载模式下平均跳数对比分析

在局部负载模式、不同的注入率下，2D Mesh 结构和正四面体裂变拓扑结构在不同的注入率下收到数据包的总数、平均跳数的实验数据如表 3 所示。由实验数据可知，该负载模式下正四面体裂变拓扑结构的平均跳数比 2D Mesh 结构约少 9.6%。两种拓扑结构在局部负载模式下的平均跳数均小于均匀负载模式下的实验值。

表 3　局部负载模式下两种拓扑结构平均跳数的对比

注入率	NoC 结构			
	2D mesh		正四面体裂变结构	
	收包总数	平均跳数	收包总数	平均跳数
0.002	1914.65	2.8193	1931.10	2.6318
0.004	3492.35	2.8206	3513.50	2.5953
0.006	3927.50	2.8614	3804.25	2.5448
0.008	3917.40	2.8245	3702.50	2.5169
0.010	3482.20	2.7009	3624.55	2.5015
0.012	3146.10	2.63091	3354.30	2.5030
0.014	3362.05	2.60823	2900.85	2.5058

3　结束语

目前，3D NoC 技术是解决片上系统集成度越来越高带来的通信瓶颈问题的重要手段。拓扑结构则是 3D NoC 研究的一个重要分支。正四面体裂变结构是一种非规则拓扑结构，本文通过仿真实验将该拓扑结构与 2D Mesh 结构在同等网络规模、相实验同条件下，主要从吞吐率、平均延迟和平均跳数三个方面与 2D Mesh 结构进行了性能对比分析。实验结果表明，在均

匀负载模式下、不同的注入率条件下，正四面体裂变拓扑结构与 2D Mesh 结构相比，平均延时更低、平均跳数更少；在局部负载模式、注入率大于 0.008 时，正四面体裂变拓扑结构与 2D Mesh 结构相比平均延时更低，在不同的注入率下，正四面体裂变拓扑结构与 2D Mesh 结构相比平均跳数更少；在吞吐量方面两种拓扑结构性能相近。实验表明：本文提出的正四面体裂变拓扑结构用于 3D NoC 拓扑结构设计是可行的。目前只是在仿真器中实现和分析了正四面体裂变拓扑结构及其性能，今后进一步的工作重点是其物理结构的设计和实现。

参考文献：

[1] Zhang Dakun, Song Guozhi, Lin Huazhou, et al. Double improved genetic algorithm and low power task mapping in 3D Networks-on-chip [J]. Journal of Computer Research and Development, 2016, 53 (4): 921-931.

[2] Dang K N, Meyer M, Okuyama Y, et al. A low-overhead softhard fault-tolerant architecture, design and management scheme for reliable high-performance many-core 3D-NoC systems [J]. Journal of Supercomputing, 2017: 1-25.

[3] Lee D, Das S, Pande P P. Performance-thermal trade-offs for a VFI-enabled 3D NoC architecture [C]// Proc of International Symposium on Quality Electronic Design. 2017: 271-276.

[4] Rezaei S H S, Modarressi M, Daneshtalab M, et al. A Three Dimensional Networks-on-Chip architecture with dynamic buffer sharing [C]// Proc of Euromicro International Conference on Parallel, Distributed, and Network-Based Processing. 2016: 771-776.

[5] Eghbal A, Yaghini P M, Bagherzadeh N, et al. Analytical fault tolerance assessment and metrics for TSV-Based 3D Network-on-Chip [J]. IEEE Trans on Computers, 2015, 64 (12): 3591-3604.

[6] 张大坤，宋国治，王莲莲，等. 三维片上网络拓扑结构研究综述 [J]. 计算机科学与探索, 2015, 9 (2): 129-164.

[7] Feero B S, Pande P P. Networks-on-chip in a three-dimensional environment: a performance evaluation [J]. IEEE Trans on Computers, 2009, 58 (1): 32-45.

[8] Jiao Jiajia, Fu Yuzhuo, Liu Ting, et al. Performance analysis and optimization for homogenous multi-core system based on 3D torus network on chip [C]// Proc of Newcas Conference. 2010: 313-316.

[9] Yin A W, Xu T C, Liljeberg P, et al. Explorations of honeycomb topologies for network-on-chip [C]// Proc of the 6th IFIP International Conference on Network and Parallel Computing. 2009: 73-79.

[10] He Xu, You Zhiqiang, Kuang Jishun. Performance analysis of three dimensional hypercube network topology [OL]. China Science and Technology Papers Online, 2011.

[11] 刘有耀，韩俊刚. 片上网络拓扑结构与通信方法研究 [D]. 西安：西安电子科技大学, 2009.

[12] 张大坤，王光兴. 基于群论的柏拉图立体着色方案三维模型构造 [J]. 软件学报, 2004, 15 (2): 292-299.

[13] Hossain H, Ahmed M, Al-Nayeem A, et al. Gpnocsim: a general purpose simulator for network-on-chip [C]// Proc of International Conference on Information and Communication Technology. 2007: 254-257.